

# Parallel Tempering in Lattice QCD with $O(a)$ -Improved Wilson Fermions

Bálint Joó, Brian Pendleton, Stephen M. Pickles and Zbigniew Sroczynski  
*Department of Physics and Astronomy, The University of Edinburgh,  
The King's Buildings, Edinburgh EH9 3JZ, Scotland*

Alan C. Irving  
*Theoretical Physics Division, Department of Mathematical Sciences  
University of Liverpool, PO Box 147, Liverpool L69 3BX, UK*

James C. Sexton  
*School of Mathematics, Trinity College and Hitachi Dublin Laboratory, Dublin 2, Ireland*

UKQCD collaboration

We present details of our investigations of the Parallel Tempering algorithm. We consider the application of action matching technology to the selection of parameters. We then present a simple model of the autocorrelations for a particular parallel tempered system. Finally we present results from applying the algorithm to lattice QCD with  $O(a)$ -improved dynamical Wilson Fermions for twin sub-ensemble systems.

12.38.Gc, 11.15.Ha, 02.70.Lq

## I. INTRODUCTION

The computational cost of lattice QCD has always been enormous. During the last few years the power of supercomputers has grown immensely but simulations with dynamical fermions are still very time consuming.

One of the most popular algorithms for dynamical fermion simulations is the Hybrid Monte Carlo (HMC) algorithm [1]. However it has been suggested that HMC is not very efficient at decorrelating some long range observables such as the topological charge [2]. On the other hand, results from the SESAM collaboration [3] indicate that HMC simulations using Wilson fermions seem to tunnel between topological sectors at an adequate rate. The results of SESAM indicate an autocorrelation time for the topological charge of about 50 HMC trajectories.

With such high computational costs it is always necessary to keep an eye open for alternative algorithms. Parallel Tempering (PT) or Exchange Monte Carlo was proposed in [4] to assist decorrelation in spin-glass systems. A lucid description of PT and related algorithms such as Simulated Tempering and their applications to spin-glass and other systems may be found in [5,6].

Recently PT has been applied to simulations of lattice QCD with staggered fermions [7] and this preliminary study indicated that the autocorrelation times for some observables were significantly improved over the normal HMC results.

In this paper we present our study of the PT algorithm using 2 flavours of degenerate  $O(a)$ -improved Wilson fermions [8] with a non-perturbatively determined coefficient [9].

PT simulates several lattice QCD ensembles concurrently, hereafter referred to as *sub-ensembles*, with different simulation parameters. PT exploits the fact that the equilibrium distributions of the configurations in individual sub-ensembles have an overlap, and occasionally tries to swap configurations between pairs of sub-ensembles, while keeping all sub-ensembles in equilibrium. This is done in such a way that the factorisation of the joint equilibrium distribution of configurations into the individual distributions for each sub-ensemble is not disturbed by the swapping.

The acceptance of these swap attempts depends on how close the sub-ensembles are to each other in parameter space. The concept of distance in parameter space is formalised in [10–12] by the machinery of action and observable matching. In theory, this technology should allow the selection of an optimal set of parameters to maximise the swap acceptance rate between the sub-ensembles.

Another possibility is to use the action matching technology to define curves in parameter space on which some observable such as  $r_0$  [13] is constant. PT can, in principle be used to simulate numerous points on such a curve in one simulation. However it must be stressed that this scenario is different from the one above. Matching observables

is not the same as matching the action [12]. Hence in this case one does not in general have as good control over the swap acceptance rate as in the situation outlined previously.

The remainder of this paper is organised as follows.

Our particular variant of the PT algorithm is described in detail in the following section, where we show that it satisfies detailed balance and present a formula for the acceptance rate of the swap attempts. We then relate this formula to the distance in parameter space as defined in the context of action matching technology.

The swapping of configurations between sub-ensembles is expected to reduce the autocorrelation times of observables within individual sub-ensembles with respect to their HMC autocorrelation times. In section III we discuss the simple case of a PT system consisting of two sub-ensembles. We suggest a model for the autocorrelation function in the PT sub-ensembles in terms of that of the original HMC ensembles.

Our simulations are discussed in Section IV and our results are presented in Section V. We show that indeed our acceptance rate formula of section II is borne out by the simulation results. We estimate the autocorrelation time of the plaquette for several swap acceptance rates and compare these estimates with the prediction of the model outlined in section III.

Our summary and conclusions are presented in Section VI.

## II. THE PARALLEL TEMPERING ALGORITHM

### Notation

Let each sub-ensemble be labelled by an index  $i$  and let the phase space of sub-ensemble  $i$  be  $\Gamma_i$ . Each sub-ensemble has an action  $\mathcal{S}_i$  which depends upon the set of parameters and the fields of the sub-ensemble.

We simulate two flavours of dynamical fermions using the standard pseudofermionic action:

$$\mathcal{S}_i = -\beta_i \mathcal{W}_\square(U) + \phi^\dagger \left( M^\dagger(\kappa_i, c_i) M(\kappa_i, c_i) \right)^{-1} \phi \quad (1)$$

where  $\mathcal{W}_\square$  is the Wilson plaquette action,  $U$  are the gauge fields,  $\phi$  are the pseudofermion fields, and  $M$  is the O(a)-improved fermion matrix with hopping parameter  $\kappa$  and clover coefficient  $c$ . In addition, for HMC algorithms we need to introduce momentum fields  $\pi_i$  and construct Hamiltonian functions  $\mathcal{H}_i = \pi_i^2 + \mathcal{S}_i$ . A state in sub-ensemble  $i$  is then represented by the triple  $a_i = (U_i, \pi_i, \phi_i)$  while the parameter set for sub-ensemble  $i$  is the triple of real numbers  $(\beta_i, \kappa_i, c_i)$ . Note that the subscript  $i$  serves only to distinguish ensembles and will be dropped when discussing a single sub-ensemble.

Each sub-ensemble has the phase space

$$\Gamma_i = \{U_i\} \otimes \{\pi_i\} \otimes \{\phi_i\}. \quad (2)$$

We note at this stage that all the  $\Gamma_i$  are identical and the only distinguishing features of individual ensembles are their parameter sets and quantities which depend upon these such as  $\mathcal{S}_i$  or  $\mathcal{H}_i$ .

A PT simulation state is thus the collection of states  $\{a_i | i = 1 \dots n\}$ , where  $n$  is the number of sub-ensembles. The overall PT phase space is the direct product of the phase spaces of the sub-ensembles

$$\Gamma_{\text{PT}} = \prod_{i=1}^n \Gamma_i. \quad (3)$$

### Detailed Balance

In a PT simulation one needs to construct a Markov process which has (joint) equilibrium probability distribution:

$$P_{\text{PT}}^{\text{eq}} = \prod_i P_i^{\text{eq}}(U, \pi, \phi) \quad (4)$$

where  $P^{\text{eq}}(U, \pi, \phi)$  is the desired equilibrium probability distribution of the individual sub-ensemble  $i$ . In our case

$$P_i^{\text{eq}}(U, \pi, \phi) = \frac{1}{Z_i} e^{-\mathcal{H}_i(U, \pi, \phi)} \quad (5)$$

$$Z_i = \int [dU] [d\pi] [d\phi] [d\phi^\dagger] e^{-\mathcal{H}_i(U, \pi, \phi)}. \quad (6)$$

Equation 4 formalises our notion of simulating ensembles independently. To be more precise, the Markov steps within any individual sub-ensemble are independent of those in the others, but the resulting sub-ensembles are not independent as they are coupled by the swapping steps. However the overall joint equilibrium distribution of the PT system is not affected by the swapping, and remains the product of the individual equilibrium distributions of the sub-ensembles.

We define two kinds of Markov transitions:

1. **Transitions within a single ensemble:** These transitions can be made with any desired Markovian update procedure that satisfies detailed balance with respect to  $P^{\text{eq}}$  for its sub-ensemble. In our case such transitions are made with HMC. We refer to the set of HMC trajectories that are performed between swaps as an HMC step.
2. **Transitions between sub-ensembles:** These transitions are used to connect the phase spaces of the sub-ensembles. Such a transition would be a proposed swap between any two sub-ensembles  $i$  and  $j$ . Let  $a$  be a configuration in sub-ensemble  $i$  and  $b$  be a configuration in sub-ensemble  $j$ . The swap transition can be denoted:

$$(a, b) \rightarrow \begin{cases} (b, a) & \text{if swap is accepted} \\ (a, b) & \text{if swap is rejected.} \end{cases} \quad (7)$$

Let us denote by  $P_s(i, j)$  the probability that the swap succeeds. The detailed balance condition is:

$$P_s(i, j) e^{-\mathcal{H}_i(a)} e^{-\mathcal{H}_j(b)} = P_s(j, i) e^{-\mathcal{H}_j(a)} e^{-\mathcal{H}_i(b)} \quad (8)$$

as the contributions from the other ensembles cancel on both sides. A suitable choice for  $P_s$  is the simple Metropolis [14] acceptance probability:

$$P_s(i, j) = \min(1, e^{-\Delta\mathcal{H}}) \quad (9)$$

where

$$\Delta\mathcal{H} = \{\mathcal{H}_j(a) + \mathcal{H}_i(b)\} - \{\mathcal{H}_i(a) + \mathcal{H}_j(b)\} \quad (10)$$

which satisfies the detailed balance condition by construction.

The required overall Markov transition should be constructed of a number of both kinds of transitions. HMC steps within all the sub-ensembles are necessary and sufficient for convergence. Transitions between sub-ensembles are not essential but without them PT would basically be the same as running several independent HMC simulations.

### Swap Acceptance Rate

Any extra decorrelation of observables in PT over and above normal HMC must necessarily come from the swapping transitions. Control of the acceptance rate for swapping transitions is therefore important. The swapping probability is determined by the energy change  $\Delta\mathcal{H}$  as in (9). The acceptance rate for Metropolis-like algorithms of this kind is easily shown to be [15]

$$\langle A \rangle = \text{erfc} \left( \frac{1}{2} \sqrt{\langle \Delta\mathcal{H} \rangle} \right). \quad (11)$$

Here  $\langle \Delta\mathcal{H} \rangle$  is the average of  $\Delta\mathcal{H}$  over all swap attempts, and  $\langle A \rangle$  is the average acceptance rate of the swap attempts.

### Action Matching

The action matching formalism outlined in [10] formalises the meaning of distance in parameter space. We review here the salient points of the discussion.

Let  $S_1[U]$  and  $S_2[U]$  be the actions of two lattice gauge theories with the same gauge configuration space, so that the partition function of each is:

$$\mathcal{Z}_i = \int [dU] \exp\{-S_i[U]\} \quad (12)$$

and the expectation of an observable  $\mathcal{O}$  in ensemble  $i$  is:

$$\langle \mathcal{O} \rangle_i = \frac{1}{Z_i} \int [dU] \mathcal{O}(U) \exp\{-S_i[U]\}. \quad (13)$$

Naturally, using actions dependent on other fields will complicate the integration measure. If one deals with pseudofermions for example, these would have to be integrated also, both in the partition function  $Z$  and in the expectation value of observables. We will not explicitly write out integrations over pseudofermions in expectation values, except cases where such an omission may lead to ambiguities.

The expectation of  $\mathcal{O}$  in the other ensemble is given to first order in a cumulant expansion by:

$$\langle \mathcal{O} \rangle_2 = \langle \mathcal{O} \rangle_1 + \langle \tilde{\mathcal{O}} \tilde{\Delta}_{12} \rangle_1 + \dots \quad (14)$$

where  $\Delta_{12} \equiv S_1 - S_2$  and  $\tilde{\mathcal{O}} \equiv \mathcal{O} - \langle \mathcal{O} \rangle$  etc.

The distance between the two actions is defined as the variance

$$d \equiv \sigma^2(\Delta_{12}) \equiv \langle \tilde{\Delta}_{12}^2 \rangle \quad (15)$$

where the expectation is to be evaluated in either sub-ensemble.

Three matching conditions have been identified:

- Match the values of observables i.e. require that  $\langle \mathcal{O} \rangle_1 = \langle \mathcal{O} \rangle_2$
- Minimise  $d$
- Maximise the acceptance in an exact algorithm for  $S_2$  constructed via accept/reject applied to configurations generated with action  $S_1$ .

It was shown in [10] that the last two conditions are equivalent to lowest order in a cumulant expansion. Under special circumstances the first condition is also equivalent to the other two to lowest order. The prescriptions differ in a calculable way at the next order.

We are now ready to make the connection between PT and the formalism of action matching. We note that the energy difference before and after a PT swap attempt is

$$\delta = \Delta \mathcal{H}. \quad (16)$$

The momentum fields cancel exactly in the Hamiltonian terms and one can deal directly with the actions:

$$\delta = S_1(U_2, \phi_2) + S_2(U_1, \phi_1) - S_1(U_1, \phi_1) - S_2(U_2, \phi_2) \quad (17)$$

Collecting the terms depending on the same fields one obtains:

$$\delta = \Delta_{12}(U_2, \phi_2) - \Delta_{12}(U_1, \phi_1). \quad (18)$$

We now identify  $\delta$  with  $-\delta$  in (3.15) in [10]. Following the analysis of [10] one may obtain the acceptance rate formula of the action matching mechanism

$$\langle A \rangle = \text{erfc} \left( \frac{1}{2} \sqrt{\sigma^2(\Delta_{12})} \right). \quad (19)$$

One can then deduce that

$$\sigma^2(\Delta_{12}) = \langle \Delta \mathcal{H} \rangle \approx \frac{1}{2} \sigma^2(\Delta \mathcal{H}) \quad (20)$$

where the second approximate equality is required to derive the acceptance rate (11).

Our PT parameters were tuned using the action matching technology to maximise the acceptance between two subensembles using the action

$$S_i = -\beta_i \mathcal{W}_\square - T_i \quad (21)$$

with

$$T_i = \text{Tr} \ln(Q_i^{-1}) \quad (22)$$

and

$$Q_i = (M^\dagger(\kappa_i)M(\kappa_i))^{-1}. \quad (23)$$

The tuning was carried out before performing the PT simulation using configurations from a preliminary HMC run at the desired reference parameter set.

However our PT simulations were carried out using the action

$$\mathcal{S}_i = -\beta_i \mathcal{W}_\square + \phi^\dagger Q_i \phi. \quad (24)$$

We have found that tuning parameters using (21) for which we had reliable technology did not optimise the swap acceptance of our simulations. The reasons for this are discussed below.

Consider first the distance  $\sigma^2$  between actions  $\mathcal{S}_i$  where the  $\mathcal{S}_i$  are as given by (21). Then

$$\Delta_{12} = \Delta\beta \mathcal{W}_\square + \Delta T \quad (25)$$

with

$$\Delta\beta = \beta_2 - \beta_1 \quad (26)$$

$$\Delta T = T_2 - T_1. \quad (27)$$

The variance of  $\Delta_{12}$  in an individual subensemble is

$$\sigma^2(\Delta_{12})_i = \langle (\Delta\beta \widetilde{\mathcal{W}_\square} + \widetilde{\Delta T})^2 \rangle_i. \quad (28)$$

One can see that for a given  $\Delta T$  one can tune  $\Delta\beta$  to minimise this variance.

However when one examines the case of the pseudofermionic action of (24) one finds that

$$\Delta_{12} = \Delta\beta \mathcal{W}_\square + \phi^\dagger (Q_1 - Q_2) \phi. \quad (29)$$

When calculating the variance of  $\Delta_{12}$  one encounters the quadratic term

$$\langle \phi^\dagger (Q_1 - Q_2) \phi \phi^\dagger (Q_1 - Q_2) \phi \rangle_i. \quad (30)$$

This term gives rise to both connected and disconnected pieces when the integration over the pseudofermion fields is carried out

$$\langle \phi^\dagger (Q_1 - Q_2) \phi \phi^\dagger (Q_1 - Q_2) \phi \rangle_i = \langle \text{Tr}^2((Q_1 - Q_2)Q_i^{-1}) \rangle_i^U + \langle \text{Tr}(Q_1 - Q_2)Q_i^{-1}(Q_1 - Q_2)Q_i^{-1} \rangle_i^U. \quad (31)$$

Here the superscript  $U$  on the expectations indicates that they are now to be carried out over the gauge fields only. Hence one finds that

$$\sigma_i^2(\Delta_{12}) = \langle (\Delta\beta \widetilde{\mathcal{W}_\square} + \text{Tr}((Q_1 - Q_2)Q_i^{-1}))^2 \rangle_i^U + \langle \text{Tr}(Q_1 - Q_2)Q_i^{-1}(Q_1 - Q_2)Q_i^{-1} \rangle_i^U \quad (32)$$

We also note that to first order in  $Q_1 - Q_2$

$$\Delta T \approx \text{Tr}((Q_1 - Q_2)Q_i^{-1}) \quad (33)$$

Comparing equations (28) and (32) it can be seen that using a pseudofermionic action gives rise to a connected piece in  $\sigma_i^2(\Delta_{12})$  which one would not get using the action of (21). This connected piece cannot be tuned away by changing  $\Delta\beta$  and it increases the distances in parameter space compared to when the action of (21) is used. If parameters are tuned using the action of (21) and the simulation is carried out using pseudofermions the acceptance rate of the swaps will not be optimised.

With hindsight it may be said that using pseudofermions was not the best choice for performing our simulations, and that the action of (21) should have been evaluated on our swap attempts to calculate  $\Delta_{12}$  instead of using the pseudofermionic action to calculate  $\Delta\mathcal{H}$ .

### III. AUTOCORRELATIONS

#### The cost of measuring observables

The gain from PT is expected to come from the swapping of configurations between sub-ensembles. This reduction in autocorrelation time is supposed to occur due to the fact that the sub-ensembles are simulated (between swaps) with independent Markov processes. However the swaps couple the ensembles and include cross correlations between them. Thus care must be taken when using results from separate subensembles together.

According to [16,17] if successive measurements of  $\mathcal{O}$  are correlated, the sample mean  $\overline{\mathcal{O}}$  is given (we use the convention of [16]) by the formula:

$$\overline{\mathcal{O}} = \langle \mathcal{O} \rangle \pm \sqrt{\frac{2\tau_{\mathcal{O}} + 1}{N} \sigma^2(\mathcal{O})}. \quad (34)$$

Here,  $\sigma^2(\mathcal{O})$  is the variance of operator  $\mathcal{O}$  given by

$$\sigma^2(\mathcal{O}) = \langle \mathcal{O}^2 \rangle - \langle \mathcal{O} \rangle^2 \quad (35)$$

and  $\tau_{\mathcal{O}}$  is the integrated autocorrelation time, defined as

$$\tau_{\mathcal{O}} = \sum_{t=1}^{\infty} C_{\mathcal{O}}(t) \quad (36)$$

and where  $C_{\mathcal{O}}(t)$  is the normalised autocorrelation function:

$$C_{\mathcal{O}}(t) = \frac{1}{\sigma^2(\mathcal{O})} \langle (\mathcal{O}(t+I) - \langle \mathcal{O} \rangle) (\mathcal{O}(I) - \langle \mathcal{O} \rangle) \rangle \quad (37)$$

and the expectation values are over all pairs of  $\mathcal{O}_i$  separated by an interval  $t$ . From now on we shall drop the subscript  $\mathcal{O}$  from these formulae except where necessary. Furthermore the term ‘autocorrelation time’ will always be used to refer to the integrated autocorrelation time.

The practical meaning of the statements above is that  $2\tau + 1$  correlated measurements of  $\mathcal{O}$  are needed in order to reduce the error in  $\overline{\mathcal{O}}$  by the same amount as if two uncorrelated measurements were used. Markov methods in general produce correlated sequences of configurations, and hence correlated sequences of measured observables. The integrated autocorrelation time  $\tau$  is therefore an important indicator of the performance of a Monte Carlo simulation that is carried out with the intention of measuring observable  $\mathcal{O}$ .

In particular, if one assumes that the autocorrelation function decays exponentially

$$C(t) = \exp\{-kt\} \quad (38)$$

with  $k > 0$ , one finds that

$$\exp\{-k\} = \frac{\tau}{\tau + 1} \quad (39)$$

which is a result we shall use later.

#### Autocorrelations in twin sub-ensemble PT

We are interested in whether or not PT will reduce the integrated autocorrelation time of an observable measured on an ensemble with some parameter set relative to the corresponding autocorrelation time of the same observable measured on an ensemble generated at the same parameters using HMC. We refer to the former of these autocorrelation times as the PT autocorrelation time and the latter as the HMC autocorrelation time.

Let us examine the situation of a PT system with two sub-ensembles. Sub-ensemble 1 has the desired parameter set, and the other sub-ensemble has its parameters chosen so as to give some acceptance rate  $\langle A \rangle$ . We assume that the HMC autocorrelation functions of both ensembles are the same. We demonstrate in section V that over the distances in parameter space for which we can use PT, and with the statistics available, we cannot differentiate between the autocorrelation times of the plaquette operator between sub-ensembles, so we regard the above assumption as reasonable.

Having made the above assumption, the changes in the autocorrelation time due to PT are now controlled solely by the number of successful swaps between the sub-ensembles. The swap probability in general depends on the

particular PT state at which the swap is attempted, but for simplicity we assume that we can replace individual swap probabilities with the overall average swap probability which is none other than the acceptance rate  $\langle A \rangle$ .

Let the HMC autocorrelation function be denoted  $C_H(t)$ , and the PT autocorrelation function of the sub-ensemble of interest be denoted  $C_{PT}(t)$ . Consider the connected autocorrelation function:

$$C_H(t) = \frac{1}{\langle \mathcal{O}^2 \rangle} \sum_{i=0}^{n-t} \mathcal{O}_{i+t} \mathcal{O}_i \quad (40)$$

where  $n$  is the number of samples of  $\mathcal{O}_i$ .

The autocorrelation function in the PT ensemble of interest can now be written as:

$$C_{PT}(t) = \frac{1}{\langle \mathcal{O}^2 \rangle} \{S_e + S_o\} \quad (41)$$

where

$$S_e = \sum_{\text{even}} \mathcal{O}_{i+t} \mathcal{O}_i \quad (42)$$

$$S_o = \sum_{\text{odd}} \mathcal{O}_{i+t} \mathcal{O}_i. \quad (43)$$

By the even sum we mean that the only terms contributing to the sum are those where an even number of swaps succeeded out of the  $t$  tried between the measurements of  $\mathcal{O}_{i+t}$  and  $\mathcal{O}_i$ .

Given some configuration in one sub-ensemble, after an odd number of successful swaps it can only be in the other one. As the HMC steps are independent in different sub-ensembles, we expect (to a first approximation) no correlation between configurations in a sub-ensemble that are separated by an odd number of swaps. Hence we assume that  $S_o$  sums to zero and we consider only the  $S_e$  term.

We then rewrite (41) as:

$$C_{PT} = P_e C_H(t) \quad (44)$$

where  $P_e$  is the probability that an even number of successful swaps occur in  $t$  trials.  $P_e$  is given by

$$P_e = \sum_i C_i^t (1 - \langle A \rangle)^{t-i} \langle A \rangle^i \quad (45)$$

where the index  $i$  runs from 0 to the largest even integer less than or equal to  $t$ ,  $i$  is even and  $C_i^t$  is the number of ways of choosing  $i$  swaps from  $t$ .

Carrying out the sum in equation (45) one finds

$$P_e = \frac{1}{2} \left\{ 1 + (1 - 2\langle A \rangle)^t \right\} \quad (46)$$

leading to the result:

$$C_{PT}(t) = \frac{1}{2} \left\{ 1 + (1 - 2\langle A \rangle)^t \right\} C_H(t). \quad (47)$$

We consider three separate cases.

- i)  $\langle \mathbf{A} \rangle = \mathbf{0}$ : In this case  $C_{PT}(t) = C_H(t)$ , which is what we expect when we do not carry out any successful swaps.
- ii)  $\mathbf{0} < \langle \mathbf{A} \rangle \leq \frac{1}{2}$ : In this case  $C_{PT} \in [\frac{1}{2}C_H, C_H)$  and we can see a reduction in the autocorrelation function of at most a factor of 2.
- iii)  $\frac{1}{2} < \langle \mathbf{A} \rangle \leq \mathbf{1}$ : In this case the term  $(1 - 2\langle A \rangle)^t$  in equation (47) becomes oscillatory. In particular if  $\langle A \rangle = 1$  (every swap succeeds) it is impossible to get an even number of successful swaps out of an odd number of trials, whereas it is a certainty for an even number of trials.

If one models the autocorrelation function by an exponential decay as in (38), it is possible to calculate the PT integrated autocorrelation time for the ensemble:

$$\tau_{PT} = \sum_1^{\infty} C_{PT}(t) \quad (48)$$

$$= \frac{1}{2}\tau_H + \frac{1}{2} \sum_1^{\infty} ((1 - 2\langle A \rangle) \exp\{-k\})^t \quad (49)$$

$$= \frac{\tau_H [1 + \langle A \rangle (\tau_H - 1)]}{1 + 2\langle A \rangle \tau_H} \quad (50)$$

where the last line follows from using (39), summing the resulting geometric series and simplifying. The ratio of  $\tau_{PT}$  to  $\tau_H$  is then:

$$\frac{\tau_{PT}}{\tau_H} = \frac{1 + \langle A \rangle (\tau_H - 1)}{1 + 2\langle A \rangle \tau_H}. \quad (51)$$

We remark on several features of the ratio in (51).

- i) When  $\langle A \rangle = 0$ , one is, in effect, carrying out two uncoupled HMC simulations and the autocorrelation times in each sub-ensemble remain the same as they would be for HMC simulations.
- ii) For a fixed  $\langle A \rangle \in (0, \frac{1}{2})$  increasing  $\tau_H$  from 0 has the effect that the ratio of (51) approaches the value of  $\frac{1}{2}$  *from above*. The closer  $\langle A \rangle$  is to  $\frac{1}{2}$ , the faster this limit is approached. If one is interested in both subensembles this is still a gain. If one of the two ensembles serves only to decorrelate the other and is not otherwise interesting (it is thrown away at the end) then one would lose over HMC as one would have done twice the work, but gained less than a factor of two.
- iii) For  $\langle A \rangle = \frac{1}{2}$  the ratio is exactly  $\frac{1}{2}$  and a breakeven is reached, in the sense that one does the work of two simulations, but in each subensemble the integrated autocorrelation is halved. This is the stage when a sub-ensemble which originally served no other purpose than to help decorrelate the other one may be thrown away without losing out.
- iv) For  $\langle A \rangle \in (\frac{1}{2}, 1]$  the ratio approaches  $\frac{1}{2}$  rapidly *from below*. In this case one clearly wins even if one is only interested in a single sub-ensemble. However the gain is not much, as for any reasonable value of  $\tau_H$  the ratio will have already approached the asymptotic limit of  $\frac{1}{2}$  to a good level of accuracy.

One can therefore win most with PT when the acceptance rate is very high, and the observable of interest has a very short autocorrelation time. In such a situation it is possible to gain more than a factor of two over the HMC autocorrelation time in each ensemble if the swap acceptance rate is greater than  $\frac{1}{2}$ . However if an observable has such a short HMC autocorrelation time, it may not be worthwhile employing PT. Parallel tempering was supposed to be used to decorrelate observables with long autocorrelation times. In a typical situation, it would be expected that the gain in each ensemble is very close to a factor of 2.

#### IV. SIMULATION DETAILS

Our PT simulations were carried out on the Cray T3E in Edinburgh. Code for performing the HMC trajectories was taken from the GHMC code written for the UKQCD Dynamical Fermions project, described in [18].

##### Program Features

The PT code ran trajectories on each sub-ensemble in *series*. Sub-ensembles were labelled from 0 to  $N - 1$ , where  $N$  was the total number of sub-ensembles. Swaps of configurations between sub-ensembles were attempted according to a boolean *plan matrix*  $M$ . If, after carrying out the HMC trajectories in sub-ensemble  $i$ , the element  $M_{ij}$  was found to contain *true*, the code would attempt to swap configurations  $j$  and  $j + 1$ . ( $j \in [0, N - 2]$ ) The default matrix had all its elements set to *false* except for the last row which had all its elements set to *true*. This way the program would perform all the HMC trajectories on all the ensembles and would then attempt a chain of pairwise swaps.

The number of HMC trajectories per sub-ensemble was controlled through an independent parameter file for each sub-ensemble. This way a sub-ensemble could be equilibrated with the GHMC code and if desired, they could easily

be taken and further simulated separately using the GHMC code. Likewise each sub-ensemble kept a separate set of log files for the plaquette and for solver statistics. The overall driver routine kept a log file of the success or failure of swap attempts and the swap energies.

### Simulation Parameters

Five PT simulations  $S1$ ,  $S2$ ,  $S3$ ,  $S4$  and  $S5$  were performed, each of which comprised two sub-ensembles. The parameters for these simulations are shown in table I. In all five simulations one sub-ensemble had parameters ( $\beta = 5.2, c = 2.0171, \kappa = .13300$ ). The parameters for the second sub-ensemble were given by action matching for  $S1$ ,  $S2$  and  $S3$ , while for  $S4$  and  $S5$  only  $\kappa$  was varied. Thus we could investigate the PT swap acceptance rate for different distances in parameter space.

We also had data from a previous HMC simulation with parameters ( $\beta = 5.2, c = 2.0171, \kappa = .13300$ ) on lattices of volume  $8^3 \times 16$  and  $8^3 \times 24$ .

The results from the reference run on the  $8^3 \times 16$  lattice were used to validate the PT code. Our PT simulations were also carried out on lattices of this size. Furthermore, it was possible to compare the autocorrelation times of the plaquette from this HMC run with the autocorrelation times of the plaquette from the first sub-ensembles of the PT runs. For the second sub-ensembles, the GHMC code was used only to achieve equilibration. Thus there is insufficient data to calculate the HMC autocorrelation times of the second sub-ensembles.

In the PT simulations each HMC step was one trajectory long. The plan matrix used was the default one described earlier. Simulations  $S1$ ,  $S2$  and  $S3$  ran for 6000 swap attempts giving 6000 trajectories for each sub-ensemble, while  $S4$  and  $S5$  ran for only 1000 swap attempts due to time constraints.

The matching procedure was performed using with the reference HMC results from  $8^3 \times 24$  lattices, using the methods outlined in [12].

### Analysis

We examined the acceptance rate as a function of the average swap energy change  $\langle \Delta \mathcal{H} \rangle$ , and of  $\Delta \kappa = \kappa_2 - \kappa_1$ , the change in the hopping parameters. We investigated the autocorrelation time of the average plaquette.

Errors in ensemble averages were estimated using the bootstrap method. Autocorrelations were estimated using the sliding window scheme of *Sokal et al.* [17].

## V. RESULTS

A summary of our results is shown in table II. We show for each simulation  $\Delta \beta = \beta_2 - \beta_1$ , the corresponding  $\Delta \kappa$ ,  $\langle \Delta \mathcal{H} \rangle$ , the acceptance rate  $\langle A \rangle$ , the integrated autocorrelation time  $\tau$  for the plaquette in sub-ensemble 1 and the autocorrelation time in sub-ensemble 1 divided by the HMC autocorrelation time,  $\tau_1 / \tau_{\text{HMC}}$ .

### Swap Acceptance Rate

Figure 1 shows the measured swap acceptance rates of the simulations. The solid line is the acceptance rate formula in (11). It can be seen that the measured results are in excellent agreement with its predictions.

### Calibration and Matching

It can be seen from table II that simulations  $S2$  and  $S3$  which had parameters given by matching the  $\text{Tr} \ln$  actions of (21) have lower acceptance rates than  $S4$  and  $S5$  for which tempering was carried out only in  $\kappa$ . We expect that this is due to the noise term of (32) and is the result of using the pseudofermion action for calculating the swap energy differences.

To see how large the effect of this noise term is, we can compare the residual variance  $\sigma^2(\Delta_{12})$  from the matching procedure [12], using the  $\text{Tr} \ln$  action with the variance as measured in our PT simulations through  $\langle \Delta \mathcal{H} \rangle$ . Note that we only have biased estimators for  $\sigma^2(\Delta_{12})$  from the matching procedure, and that we have calculated the residual variance estimate only for  $\Delta \kappa = .0005$ .

Table III contains our comparison of the  $\text{Tr} \ln$  matching predictions and pseudofermionic measurements for simulation  $S3$ . We can see in column 2, our biased estimate of the residual variance on matching and in column 4 the corresponding predicted acceptance rate. In column 3 we see the actual variance as measured in the simulation and in column 6 the corresponding measured acceptance rate. We expect the difference in the variances to be due to the four point term in equation (32). We can therefore numerically estimate the four point term to be

$$\langle \text{Tr}(Q_2 - Q_1)Q_i^{-1}(Q_2 - Q_1)Q_i^{-1} \rangle_i^U = 6.6(2) \quad (52)$$

for simulation  $S3$ .

Note that if during our swap acceptance steps, we would discard the pseudofermion fields, and calculate the energy change using the  $\text{Tr} \ln$  action by the methods outlined in [12], we would suffer a workload hit, but would expect

an accept rate of around 48% in the case of simulation *S3*. Thus using pseudofermions was a poor way to proceed originally. However as the action difference scales like the lattice volume, going to larger lattices would effectively cancel all the gain one could obtain by using the  $\text{Tr} \ln$  action to evaluate the swap action/energy difference.

### Autocorrelation Times and Efficiency

The autocorrelation times of the plaquette operator on the sub-ensembles with parameter  $\kappa = .1330$  are shown in column 5 of table II. We also show for comparison the autocorrelation time estimated from our independent HMC run at the same parameter set. In table IV we gather some estimates of the integrated autocorrelation time of the plaquette for some independent HMC runs at similar parameters to our PT runs. It can be seen that the HMC autocorrelation times agree with each other within estimated errors, justifying the assumptions of our model of section II.

Figure 2 shows the ratio of PT to HMC autocorrelation times. The errors on the ratios were obtained by simple error combination. The line superimposed on the data in figure 2 is the prediction of the model in section II (c.f. equation 51). It can be seen that it is not inconsistent with the data.

## VI. SUMMARY AND CONCLUSIONS

In this paper we presented our study of the Parallel Tempering algorithm applied to lattice QCD with  $O(a)$ -improved Wilson fermions. We showed how the algorithm satisfies detailed balance, and gave a formula for the swap acceptance rate in terms of the swap energy change  $\Delta\mathcal{H}$ . We highlighted the connection of parallel tempering with the technology of action matching. We presented and discussed a simple model of autocorrelations in a twin sub-ensemble PT system, and found that the algorithm is unlikely to improve autocorrelation times by more than a factor of two for such a system. We verified our simple model assumptions by gathering autocorrelation time data from previous simulations.

We carried out a numerical study where we verified the acceptance formula and the predictions of the autocorrelation model within statistical errors. We also obtained information on how the acceptance rate of the algorithm falls with increasing  $\Delta\kappa$ .

We found that using the pseudofermions from HMC on the swap attempt is a poor way to proceed if the parameters are matched for the  $\text{Tr} \ln$  action. We have shown analytically that there is an extra noise term in the definition of the distance between actions when pseudofermions are used. We have attempted to estimate the size of this noise term numerically.

We conclude that Parallel Tempering does not seem to give any real gain over HMC at the present time. We were unable to use PT to simulate sub-ensembles sufficiently far apart in parameter space. The acceptance rate drops too quickly with  $\Delta\kappa$ . This situation could be alleviated somewhat if the swap action/energy differences were to be calculated using the  $\text{Tr} \ln$  action, for simulations with parameters matched with that action. However in the end the real problem is that the swap action/energy change scales with the volume for a fixed kappa, and that when employing the PT algorithm on a realistic sized (eg  $16^3 \times 32$ ) lattice, the scaling of the swap energy change would lower the acceptance rate and lose all that could be gained by using the  $\text{Tr} \ln$  action.

Thus we could not take advantage of the fact that in one region of parameter space autocorrelation times are short while in another they are long. With our parameter values, the HMC autocorrelation times of our sub-ensembles are the same within experimental errors and the predictions of our model apply. A chain of sub-ensembles that would span the required distance in parameter space can be constructed, but would take an unfeasibly large number of sub-ensembles for lattices of interesting size.

## ACKNOWLEDGMENTS

We gratefully acknowledge support from PPARC grant no GR/L22744, and EPSRC for funding under grant number GR/K41663. We also wish to thank Tony Kennedy and Stephen Booth for helpful discussions.

- 
- [1] S. Duane, A. D. Kennedy, B. J. Pendleton, D. Roweth, *Phys. Lett.* **B195** (1987) 216-222.
  - [2] B. Allés, G. Boyd, M. D’Elia, A. Di Giacomo, E. Vicari, *Phys.Lett.* **B389** (1996) 107-111, *hep-lat/9600749*

- [3] B. Allés, G. Bali, M. D’Elia, A. Di Giacomo, N. Eicker, S. Güesken, H. Höeber, Th. Lippert, K. Schilling, A. Spitz, T. Struckmann, P. Ueberholz, J. Viehoff, *hep-lat/9803008*.
- [4] K. Hukushima, J. Nemoto, *cond-mat/9512035*.
- [5] E. Marinari, *cond-mat/9612010*
- [6] E. Marinari, G. Parisi, J. Ruiz-Lorenzo, *Spin Glasses and Random Fields*, edited by P. Young, *cond-mat/9701016*
- [7] G. Boyd, *Nucl. Phys. (Proc. Suppl.)* **60A** (1998) 341-344, *hep-lat/9712012*
- [8] B. Sheikholeslami and R. Wohlert, *Nucl. Phys.* **B259** (1985) 572.
- [9] K. Jansen and R. Sommer, *Nucl. Phys. B. (Proc. Suppl.)* **63A-C** (1998) 853-855, *hep-lat/9709022*
- [10] A. C. Irving and J. C. Sexton, *Phys. Rev.* **D55** (1997) 5456.
- [11] A. C. Irving, J. C. Sexton and E. Cahill, *Nucl. Phys. B. (Proc. Suppl.)* **63A-C** (1998) 967
- [12] A. C. Irving, J. C. Sexton, E. Cahill, J. Garden, B. Joó, S. M. Pickles and Z. Sroczynski, UKQCD Collaboration, *hep-lat/9807015*
- [13] R. Sommer, *Nucl. Phys.* **B411** (1994) 839.
- [14] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller, *J. Chem. Phys* **21** (1953) 1087
- [15] S. Gupta, A. Irbäck, F. Karsch and B. Petersson. *Phys. Lett.* **B242** (1990) 437
- [16] I. Horváth, and A. D. Kennedy, *Nucl. Phys.* **B510** (1998) 367-400, *hep-lat 9708024*
- [17] N. Madras, A. D. Sokal, *J. Stat. Phys.* **50** (1988) 109-186
- [18] Z. Sroczynski, S. M. Pickles and S. P. Booth, UKQCD Collaboration, *Nucl. Phys. B. (Proc. Suppl.)* **63A-C** (1998) 949-951.

TABLE I. Simulation parameters used for twin ensemble runs and the reference HMC run

[h] Simulation	$(\beta_1, c_1, \kappa_1)$	$(\beta_2, c_2, \kappa_2)$
HMC	(5.2, 2.0171, 0.133)	
S1	(5.2, 2.0171, 0.133)	(5.2060, 2.01002, 0.13280)
S2	(5.2, 2.0171, 0.133)	(5.2105, 2.00471, 0.13265)
S3	(5.2, 2.0171, 0.133)	(5.2150, 1.99940, 0.13250)
S4	(5.2, 2.0171, 0.133)	(5.2, 2.0171, 0.13280)
S5	(5.2, 2.0171, 0.133)	(5.2, 2.0171, 0.13265)

TABLE II. Results from the PT simulations showing the appropriate results from HMC for comparison

Simulation	$\Delta\beta(\times 10^{-3})$	$\Delta\kappa(\times 10^{-4})$	$\langle\Delta\mathcal{H}\rangle$	$\langle A\rangle$	$\tau_1$	$\tau_1/\tau_{\text{HMC}}$
HMC					26(6)	1
S1	6	-2.0	1.23(2)	0.43(1)	12(3)	0.5(2)
S2	10.5	-3.5	3.76(4)	0.17(1)	19(4)	0.7(2)
S3	15	-7.5	7.64(6)	0.051(2)	24(6)	0.9(3)
S4	0	-2.0	0.91(4)	0.49(1)	9(4)	0.3(2)
S5	0	-3.5	2.29(7)	0.26(2)	18(10)	0.7(4)

TABLE III. Comparison of Tr ln matching and acceptance with pseudofermionic acceptance

Simulation	$\sigma^2(\Delta_{12})_{\text{Tr ln}}$	$\sigma^2(\Delta)_{\text{p.f}} = \langle\Delta\mathcal{H}\rangle$	$\langle A\rangle_{\text{Tr ln}}$	$\langle A\rangle_{\text{p.f}}$
S3	1.02(20)	7.64(6)	0.48(5)	0.051(2)

TABLE IV. The integrated autocorrelation times of some other simulations.

$\beta$	$c$	$\kappa$	$\tau_{\text{HMC}}$
5.2	1.99	.1335	18(8)
5.2	2.0171	.1330	26(6)
5.232	1.98	.1335	20(6)

## Swap Acceptance Rate v.s. $\langle \Delta H \rangle$

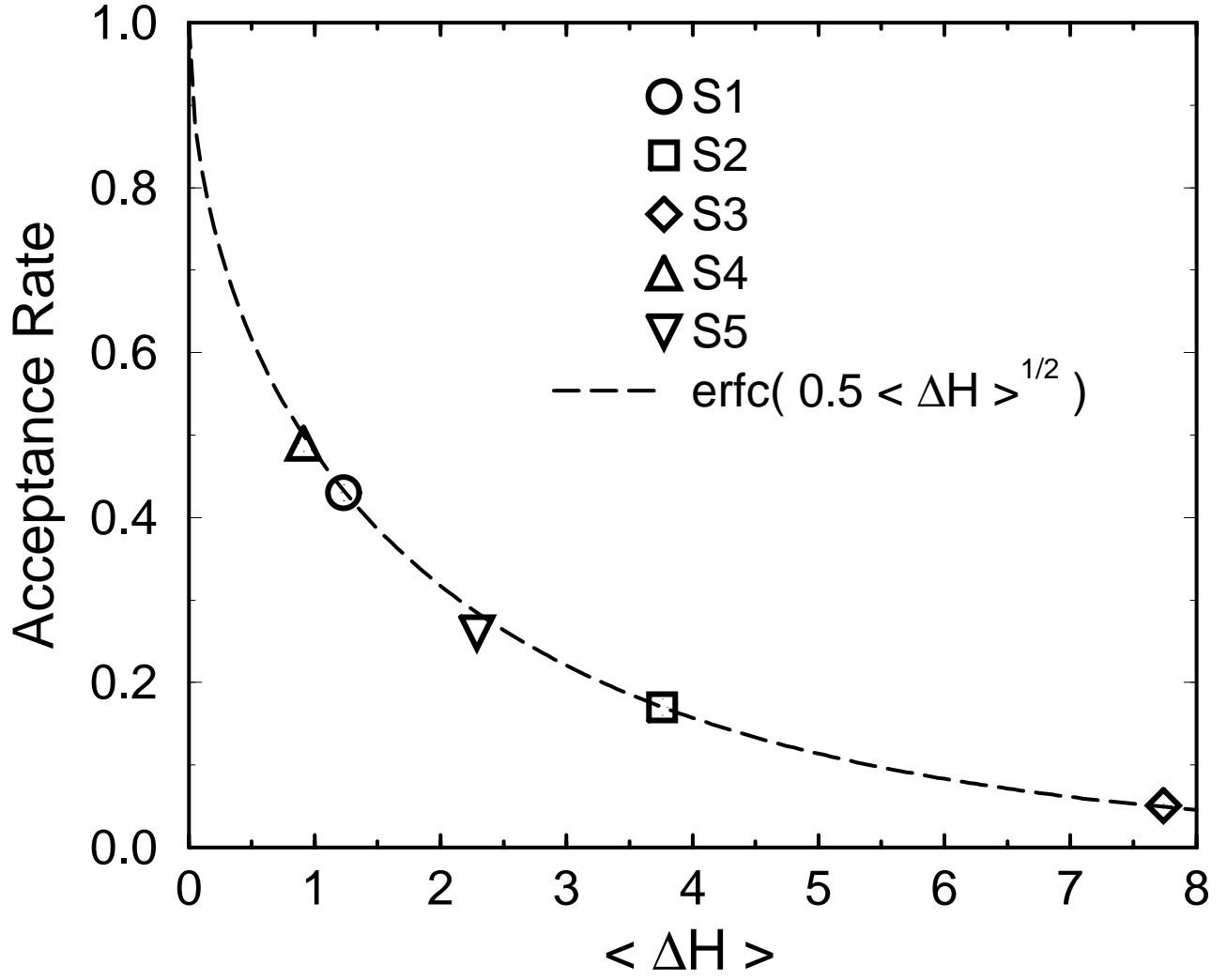


FIG. 1. Acceptance rate against  $\langle \Delta H \rangle$ . Error bars are smaller than the symbols

# Comparison of Autocorrelation Times

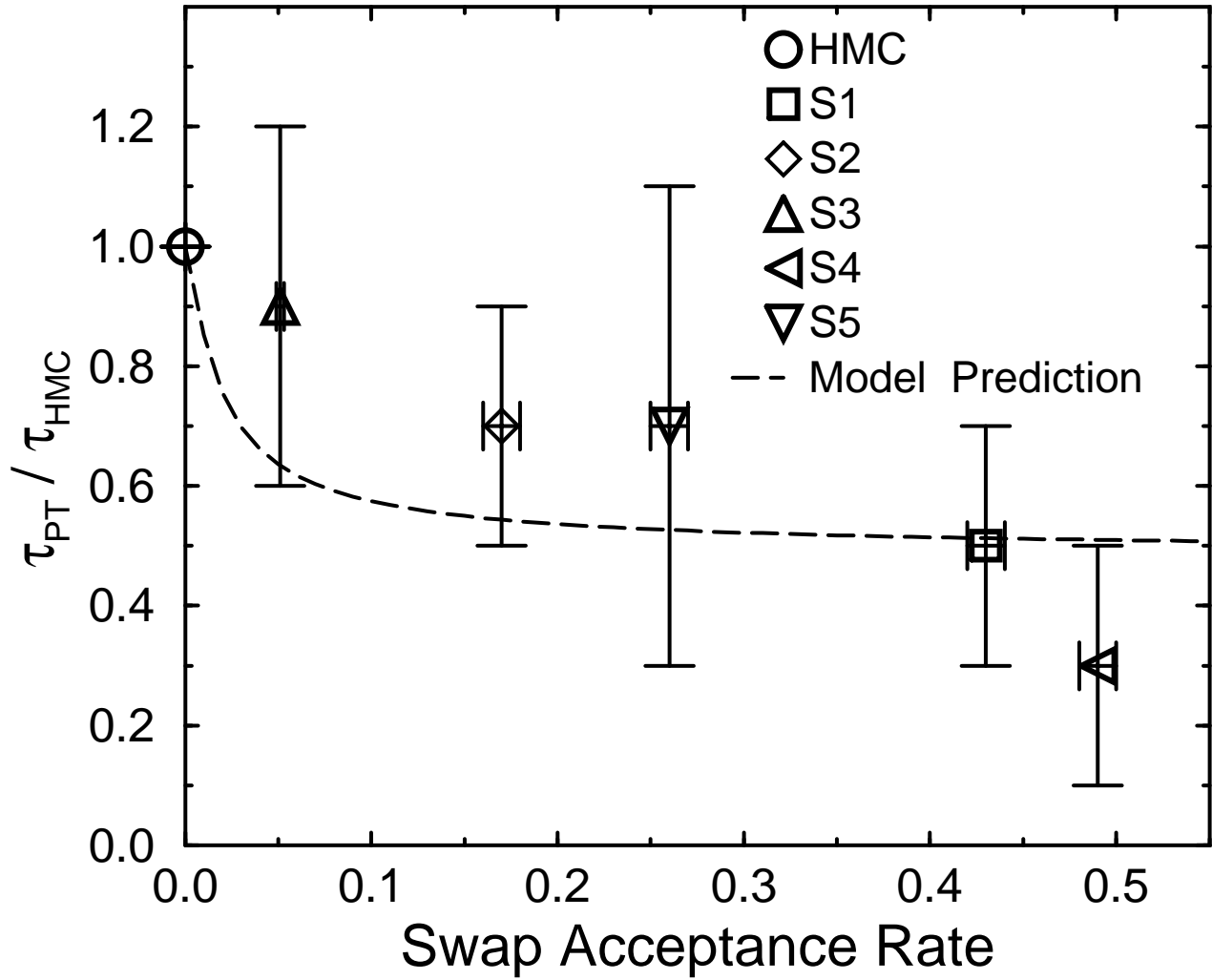


FIG. 2. Integrated autocorrelation times for the plaquette normalised by that from GHMC simulations